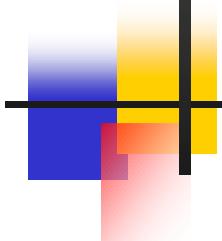


Measurement of physical quantities in the Bayesian framework using neural networks



***Advanced Statistical Techniques
in Particle Physics***

Durham, 18-22 March 2002

Marcin Wolter

Tufts University, Boston

Institute of Nuclear Physics, Krakow

Main goal

- **Apply the Bayesian method to the particle mass reconstruction.**
- *No explicit knowledge of the functional dependence of the mass estimate is needed.*
- *Bayesian method makes, in principle, the optimal use of the information supplied.*



Rev. Thomas Bayes



Thomas Bayes (b. 1702, London - d. 1761, Tunbridge Wells, Kent), mathematician who first used probability inductively and established a mathematical basis for probability inference (a means of calculating, from the number of times an event has not occurred, the probability that it will occur in future trials). He set down his findings on probability in "Essay Towards Solving a Problem in the Doctrine of Chances" (1763), published posthumously in the *Philosophical Transactions of the Royal Society of London*.

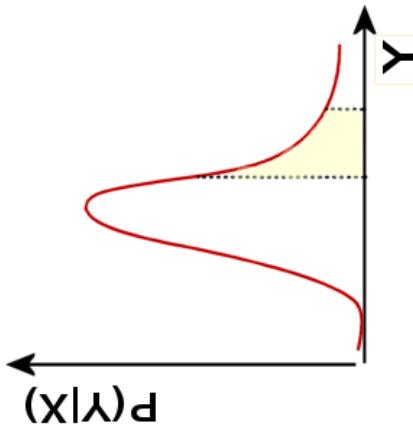
Article on WEB: <http://www.uv.es/valencia7> (J. Bernardo)

Book: D.S. Sivia "Data Analysis - a Bayesian Tutorial"



Statistics of Bayes & Laplace

- Probability is a *degree-of-belief* or plausibility (not defined by frequency of occurrence).
- To measure the parameter Y having data X we look for the conditional probability $P(Y/X)$.
- In the analysis we do not need to know the functional dependence $Y=f(X)$, it is enough to know $P(X/Y)$ and than obtain $P(Y/X)$ from the



Bayes theorem:

$$P(\text{hypothesis}/\text{data}) \propto P(\text{data}/\text{hypothesis}) \times P(\text{hypothesis})$$



Data analysis in a Bayesian way

Measuring the particle mass (an example).

- For each mass M_i find $P(x/M_i)$, where x is a vector of measured variables.
- Bayes theorem:
$$P(hypothesis/data) \propto P(data/hypothesis) \times P(hypothesis)$$
therefore:
$$P_n(M_i | x_1, \dots, x_n) \propto \left(\prod_{l=1}^n P(x_l | M_i) \right) \times P(M_i)$$
for a sample of n events.
 - The mode of $P(M/x_1 \dots x_n)$ distribution is the mass estimate.

Introduction

- Reconstruct the Higgs mass from the process $H \rightarrow b\bar{b} \rightarrow 2\text{jets}$ (PGS simulation) using Bayesian approach (i.e. probability distribution as a function of Higgs mass calculated for every event).
- Use information normally used to calculate invariant mass: E_1, E_2 and $\cos(\alpha)$.
- Find correct Higgs mass for not calibrated jet energies.
- No background present (neither physical nor combinatorial).



Monte Carlo simulation

- All events were generated using CompHEP version 33.23, hadronized with Pythia version 6.158 and event book-keeping was done using STDHEP version 4.09.
- The detector simulation used PGS version 3.0 (Pretty Good Simulation, fast simulation package developed by John Conway et al.) with the following parameters:

- number of eta cells in calorimeter: 84
- number of phi cells in calorimeter: 64
- width of each cell in eta: 0.10
- width of each cell in phi: 0.10
- em calorimeter resolution: 0.20
- hadron calorimeter resolution: 0.80
- trigger MET resolution: 0.20
- calorimeter cell edge crack fraction: 0.05
- cluster seed tower threshold: 3.00
- shoulder tower threshold: 0.50
- cluster finder cone size: 0.70
- outer radius of tracking (m): 1.00
- magnetic field (T): 1.40
- sagitta resolution (m): 0.00004
- track efficiency: 0.98
- minimum track pt: 0.30
- tracking eta coverage: 2.00

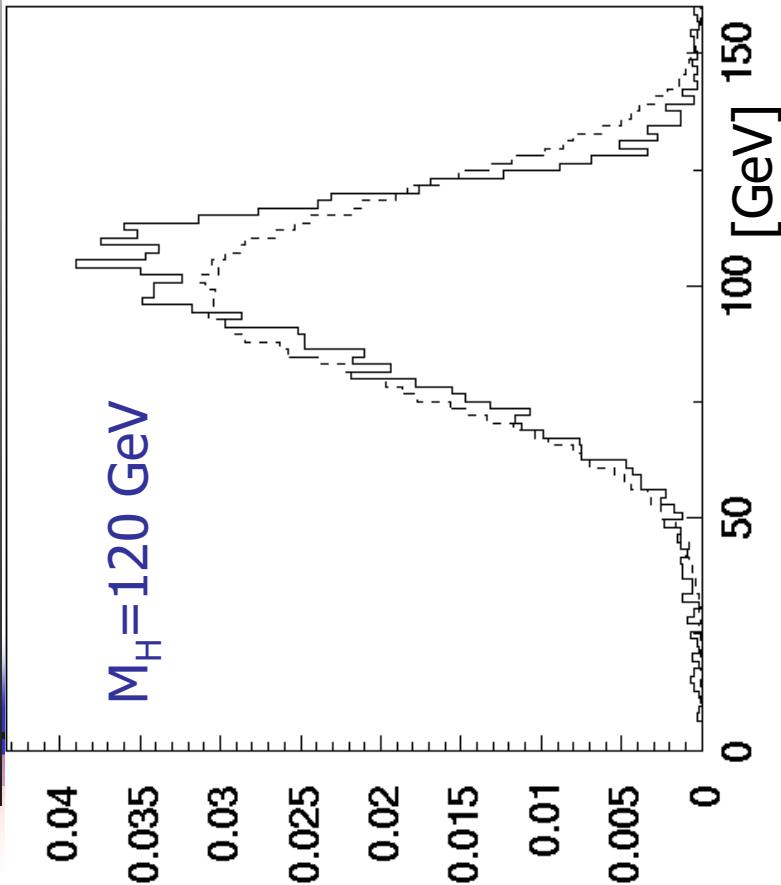
Method description

- Generate MC samples with different Higgs masses (95 GeV – 155 GeV, 5 GeV increments).
- Find a E_1, E_2 and $\cos(\alpha)$ distribution for every Higgs mass.
- Fit the distributions using neural network.
- Normalize all distributions.
- For every event find $P(M_H/x) \propto P(x/M_H) \times P(M_H)$. For n events:

$$P_n(M_H | x_1, \dots, x_n) \propto \left(\prod_{i=1}^n P(x_i | M_H) \right) \times P(M_H)$$

- Weighted average of $P_n(M_H/x)$ is chosen as the Higgs mass estimate.

Fitting using a neural network*



- Add a sample of events with flat distribution in E_1 , E_2 and $\cos(\alpha)$ to MC sample.
- Use it as a "background" for NN training (3 inputs, 1 output) and MC events as a "signal".
- Probability: $P(x, m) \propto \frac{NN_{out}}{(1 - NN_{out})}$
- **Advantages:** unbinned fit, no analytical function needed, trained NN easy to convert into a fast subprogram.

Fit quality could be improved by generating bigger MC samples – here only 2300 events used for 3D fit.

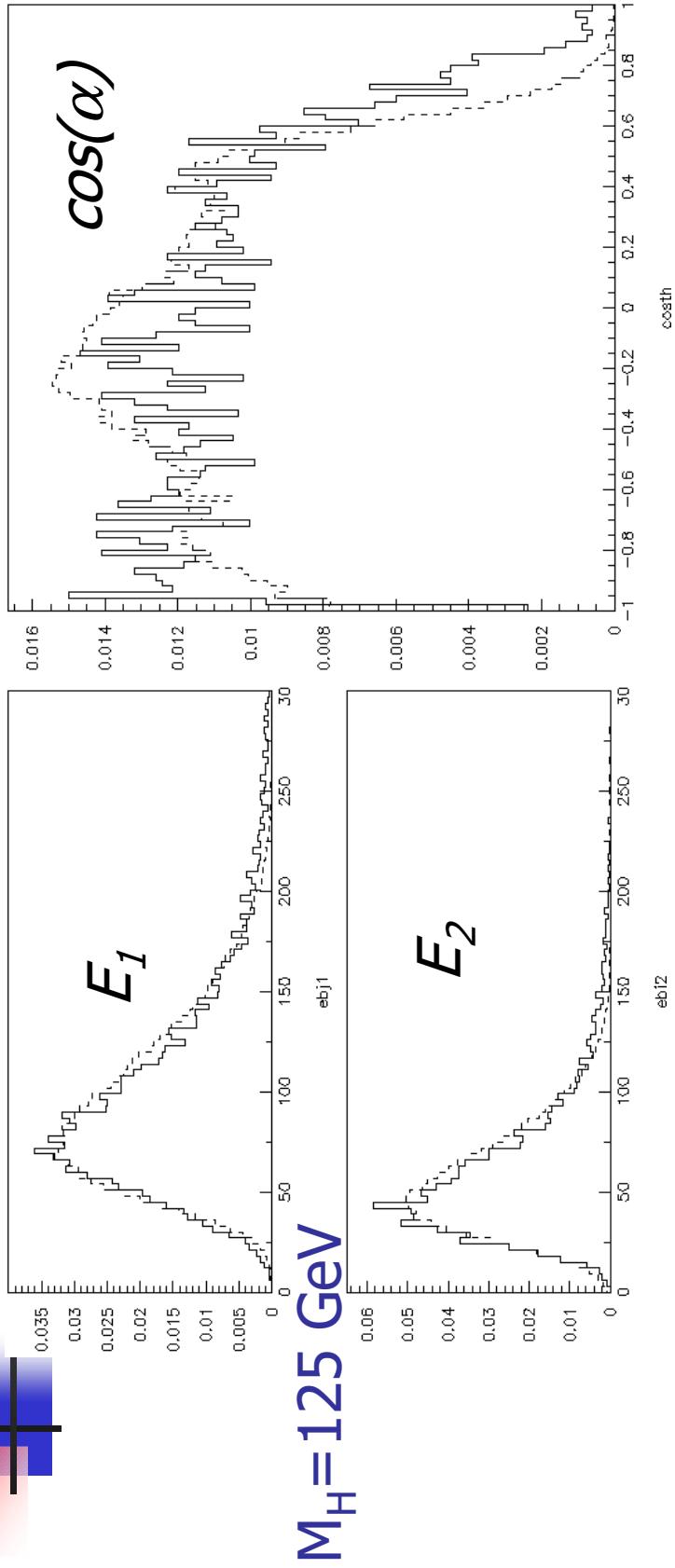
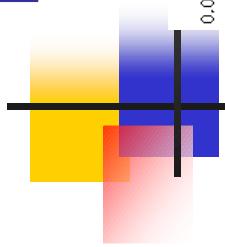
Invariant mass (solid line) compared with the invariant mass from the fitted function (dashed line).

Durham, 18-28 March 2002

M. Wolter, Measurement of physical quantities...

*L.Garrido, A.Juste, Comp. Phys. Com. 115 (1998)

Fit quality



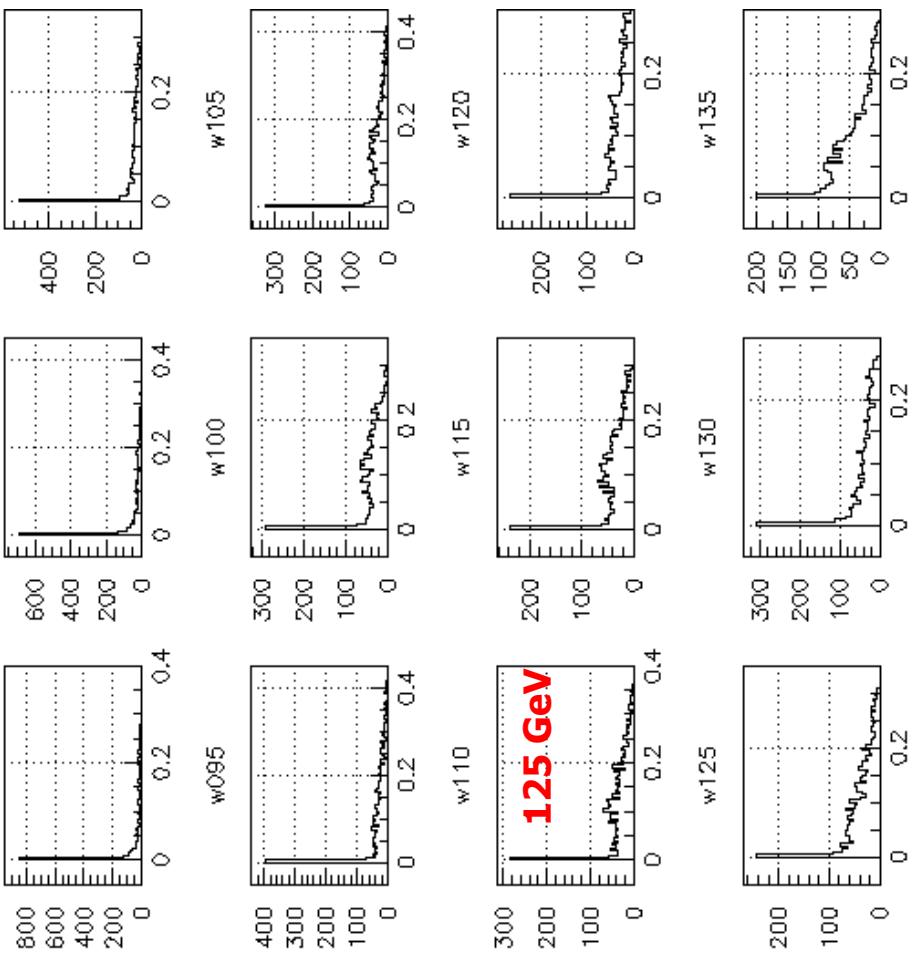
Distributions of E_1 , E_2 and $\cos(\alpha)$ for $M_H = 125 \text{ GeV}$ sample (solid line) compared with the distributions obtained from the fitted function.

Probability distributions

- Correct Higgs mass reconstruction - modes of the probability distributions (~ 2000 events) at correct Higgs masses.
 - Logarithm $\ln(P)$ on the Y-axis.
-
- The figure consists of four subplots arranged in a 2x2 grid. Each subplot shows a histogram of the logarithm of the probability $\ln(P)$ on the Y-axis (ranging from -8000 to -5000) against mass in GeV on the X-axis (ranging from 100 to 140). The subplots are labeled with their respective Higgs masses: 110 GeV, 120 GeV, 130 GeV, and 140 GeV. In each plot, there is a prominent peak at a specific mass value, indicated by a vertical blue line. The peaks are located at approximately 110 GeV, 120 GeV, 130 GeV, and 140 GeV respectively.



Single event probabilities



Distributions of single events probabilities for $M_H = 125$ Gev sample.

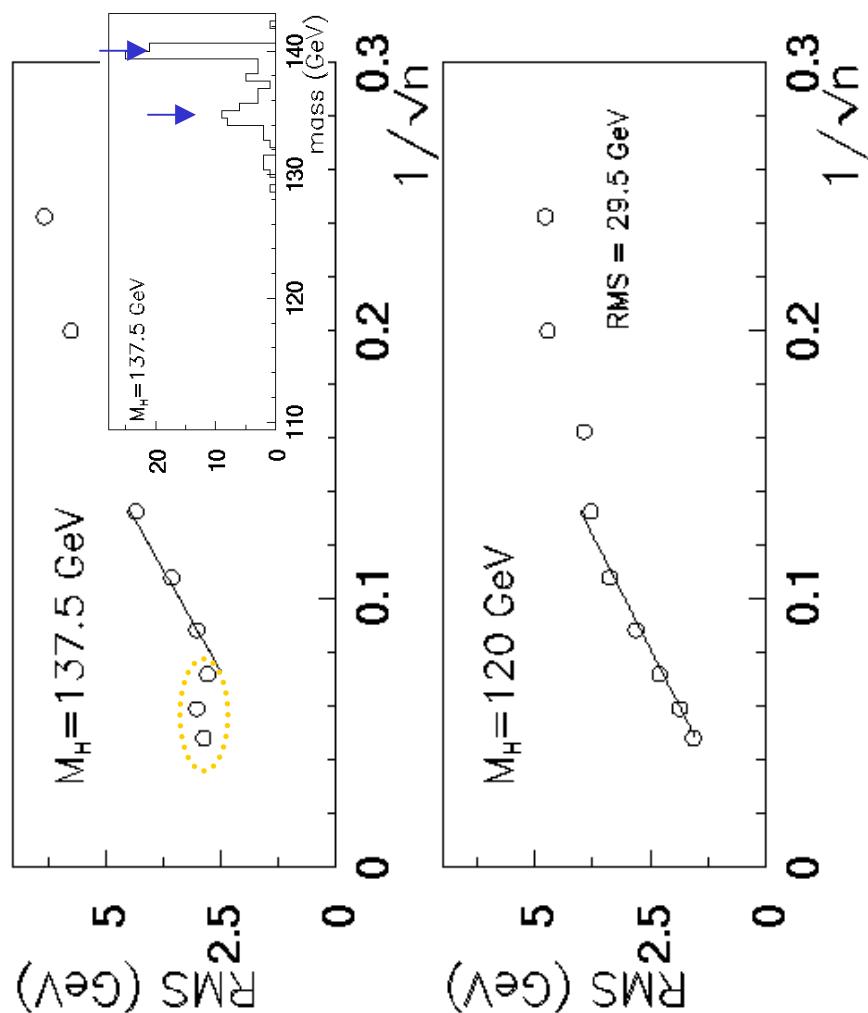
Highest probabilities for 125 GeV, as expected.

Particle mass reconstruction

- Reconstructed mass as a function of the number of events in a sample.
 - Mass properly reconstructed (*events with $M_H=137.5\text{ GeV}$ were not used for training*).
 - "Edge effect" due to the weighted average taken over a finite Higgs mass range.
-
- $M_H=137.5\text{ GeV}$
- | no. of events | mass (GeV) |
|---------------|------------|
| 0 | 137.5 |
| 10 | 137.5 |
| 20 | 137.5 |
| 30 | 137.5 |
| 40 | 137.5 |
| 50 | 137.5 |
| 60 | 137.5 |
| 70 | 137.5 |
| 80 | 137.5 |
| 90 | 137.5 |
| 100 | 137.5 |
| 110 | 137.5 |
| 120 | 137.5 |
| 130 | 137.5 |
| 140 | 137.5 |
| 150 | 137.5 |
| 160 | 137.5 |
| 170 | 137.5 |
| 180 | 137.5 |
| 190 | 137.5 |
| 200 | 137.5 |
| 210 | 137.5 |
| 220 | 137.5 |
| 230 | 137.5 |
| 240 | 137.5 |
| 250 | 137.5 |
| 260 | 137.5 |
| 270 | 137.5 |
| 280 | 137.5 |
| 290 | 137.5 |
| 300 | 137.5 |

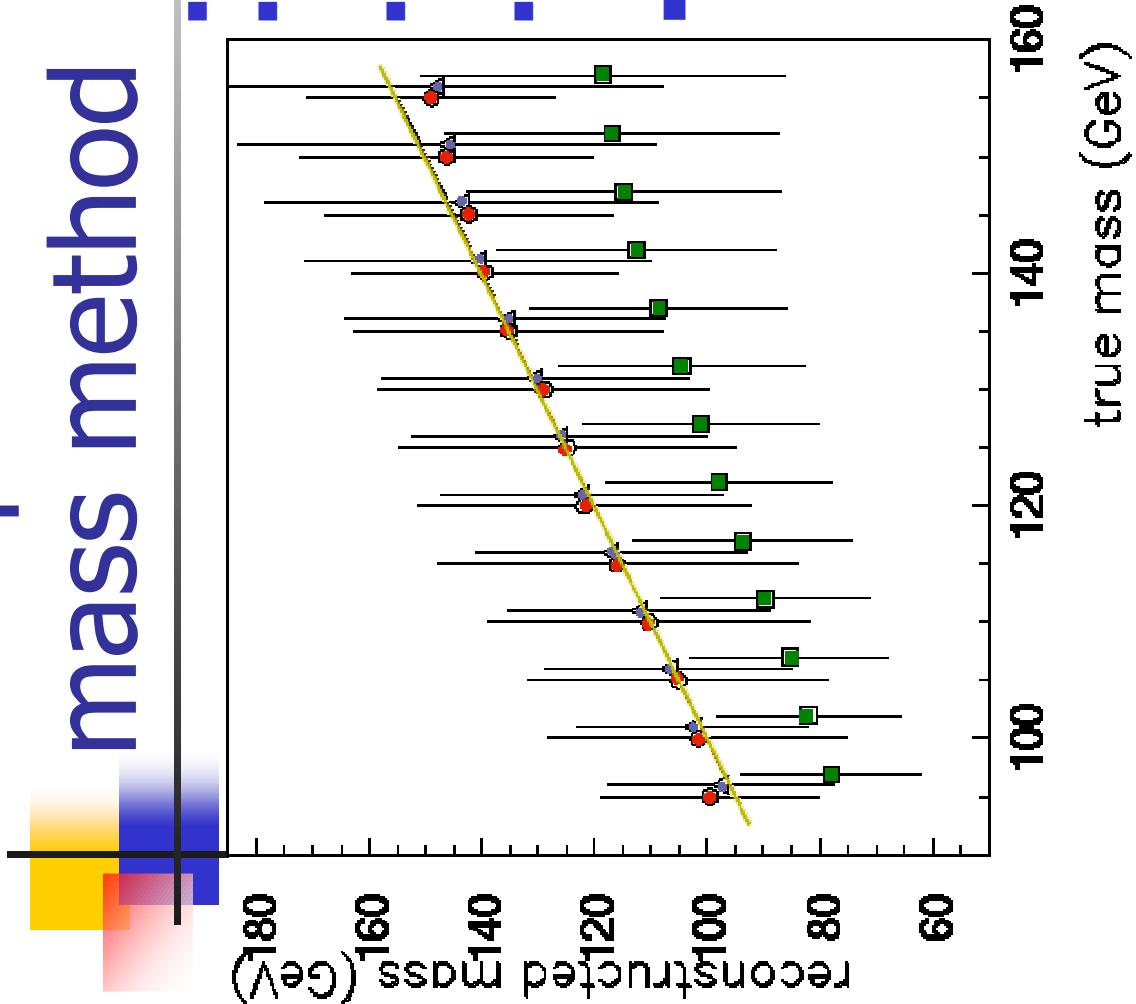
Resolution estimation

- Finding RMS of the estimated mass distribution for various sizes N of event samples.
- Obtaining “single event” RMS from the fit:



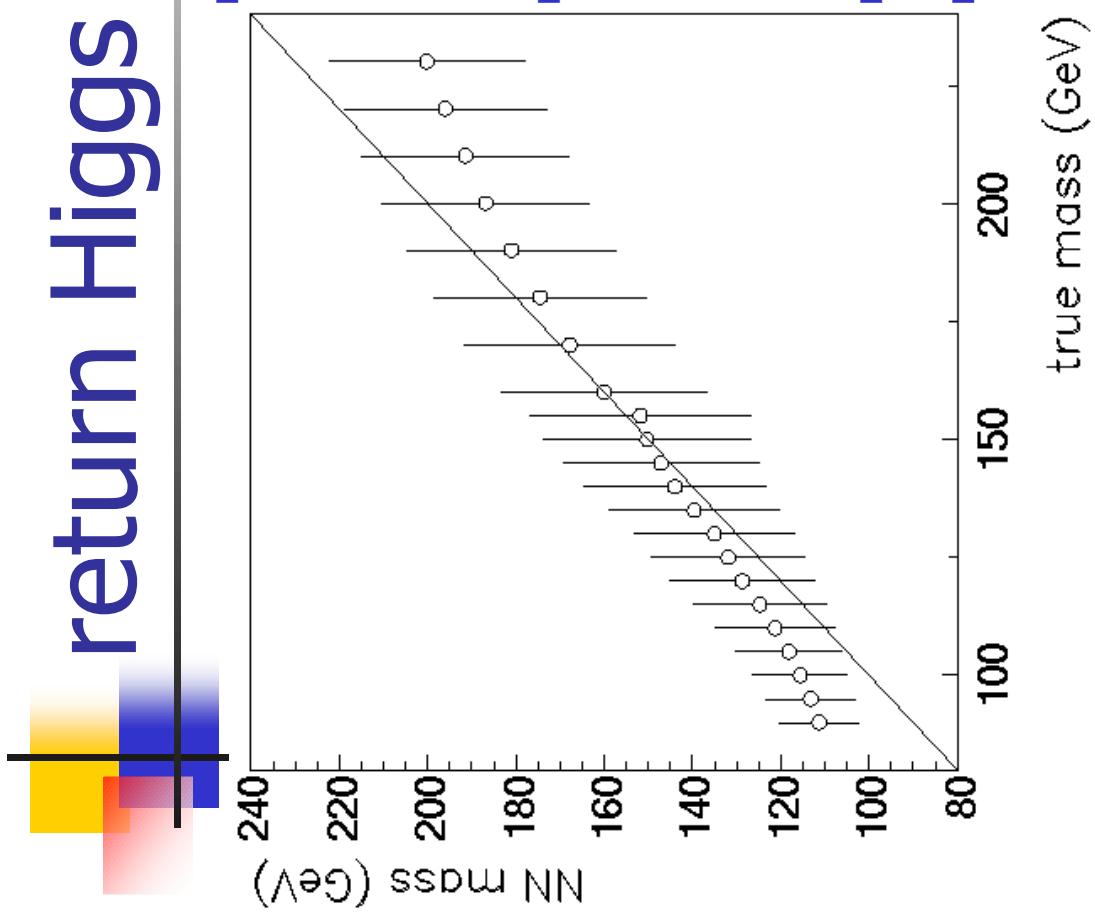
- Two peaks for $M_H = 137.5 \text{ GeV}$ – the algorithm tends to return one of the neighbor masses. Effect could be minimized by generating more MC samples.

Comparison with invariant mass method



- Bayesian approach gives results very similar to the use of invariant mass, but no correlations encoded in the invariant mass formula are used.

Neural network trained to return Higgs mass directly

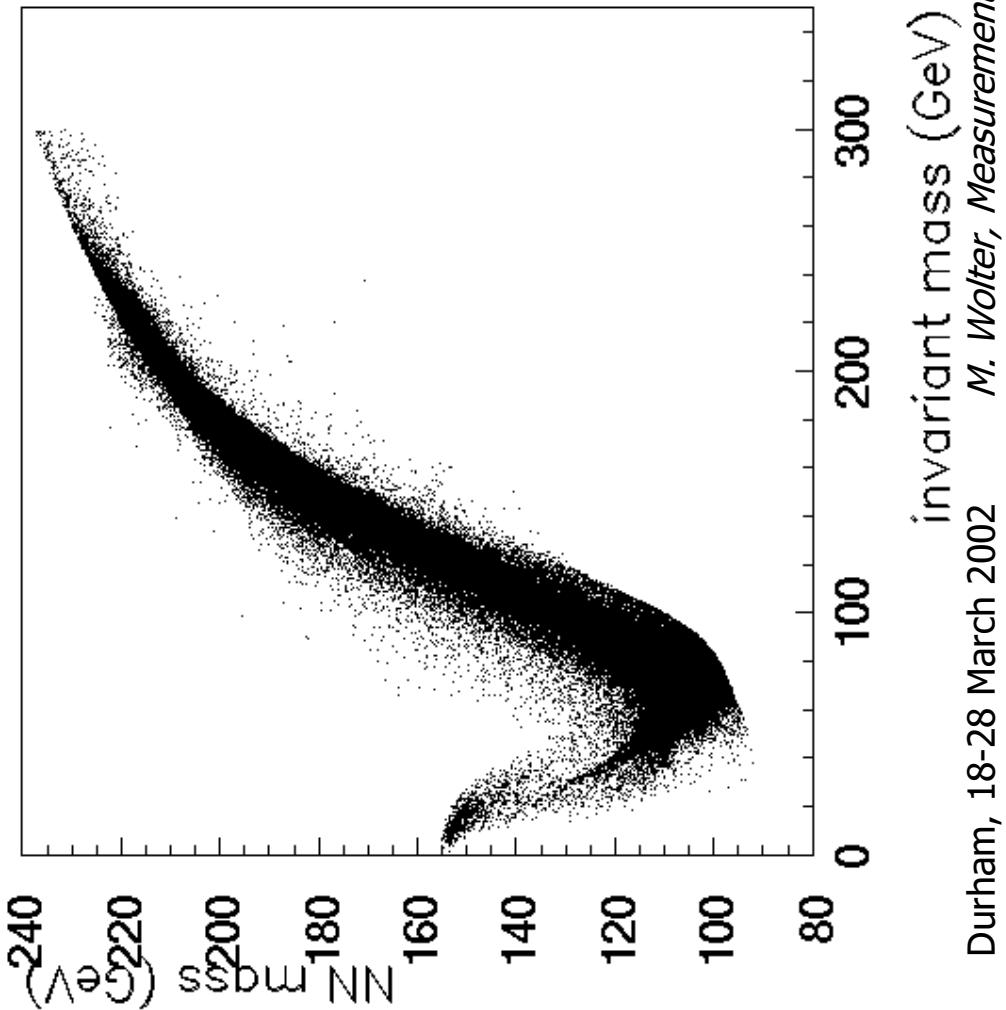


- Single NN trained on MC samples $M_H=90$ GeV up to $M_H=230$ GeV. It returns Higgs mass directly.
- NN minimizes χ^2 , therefore reconstructed masses are shifted towards the middle of the spectrum.
- RMS from 10 GeV up to 23GeV
- Could become linear if MC generated with wider Higgs mass range.



Network returning Higgs mass

Higgs mass reconstructed by a neural network vs. the standard invariant mass
– the “mass shifting” effect due to the χ^2 minimization is clearly visible.



Conclusions

- Bayesian method works well and can be applied without knowing the functional dependence of particle properties on the measured quantities.
- No more information can be extracted from the E_1, E_2 and $\cos(\alpha)$ beyond that encoded in the invariant mass.
- By adding additional variables it may be possible to improve the power of the Bayesian method.
- Neural network is a good fitting tool – unbinned multidimensional fits.